國立政治大學統計學系博士學位論文

指導教授：劉惠美 博士

偏常態因子信用組合下之效率估計值模擬

Efficient Simulation in Credit Portfolio with Skew Normal Factor

研究生：林永忠 撰

中華民國一百零一年七月

中文摘要

在因子模型下，損失分配函數的估算取決於混合型聯合違約分配。蒙地卡羅是一個經常使用的計算工具。然而，一般蒙地卡羅模擬是一個不具有效率的方法，特別是在稀有事件與複雜的債務違約模型的情形下，因此，找尋可以增進效率的方法變成了一件迫切的事。

對於這樣的問題，重點採樣法似乎是一個可以採用且吸引人的方法。透過改變抽樣的機率測度，重點採樣法使估計量變得更有效率，尤其是針對相對複雜的模型。因此，我們將應用重點採樣法來估計偏常態關聯結構模型的尾部機率。這篇論文包含兩個部分。Ⅰ：應用指數扭轉法---一個經常使用且為較佳的終點採樣技巧---於條件機率。然而，這樣的程序無法確保所得的估計量有足夠的變異縮減。此結果指出，對於因子在選擇重點採樣上，我們需要更進一步的考慮。Ⅱ：進一步應用重點採樣法於因子；在這樣的問題上，已經有相當多的方法在文獻中被提出。在這些文獻中，重點採樣的方法可大略區分成兩種策略。第一種策略主要在選擇一個最好的位移。最佳的位移值可透過操作不同的估計法來求得，這樣的策略出現在 Glasserman 等(1999)或 Glasserman 與 Li (2005)。

第二種策略則如同在 Capriotti (2008)中的一樣，則是考慮擁有許多參數的因子密度函數作為重點採樣的候選分配。透過解出非線性優化問題，就可確立一個未受限於位移的重點採樣分配。不過，這樣的方法在尋找最佳的參數當中，很容易引起另一個效率上的問題。為了要讓此法有效率，就必須在使用此法前，對參數的穩健估計上，投入更多的工作，這將造成問題更行複雜。

本文中，我們說明了另一種簡單且具有彈性的策略。這裡，我們所提的演算法不受限在如同 Gaussian 模型下決定最佳位移的作法，也不受限於因子分配函數參數的估計。透過 Chiang, Yueh 與 Hsie (2007)文章中的主要概念，我們提供了重點採樣密度函數一個合理的推估並且找出了一個不同於使用隨機近似的演算法來

加速模擬的進行。

最後，我們提供了一些單因子的理論的證明。對於多因子模型，我們也因此有了一個較有效率的估計演算法。我們利用一些數值結果來凸顯此法在效率上，是遠優於蒙地卡羅模擬。

關鍵字：蒙地卡羅模擬；重點採樣法；信用風險組合；變異縮減。

ABSTRACT

Under a factor model, computation of the loss density function relies on the estimates of some mixture of the joint default probability and joint survival probability. Monte Carlo simulation is among the most widely used computational tools in such estimation. Nevertheless, general Monte Carlo simulation is an ineffective simulation approach, in particular for rare event aspect and complex dependence between defaults of multiple obligors. So a method to increase efficiency of estimation is necessary.

Importance sampling (IS) seems to be an attractive method to address this problem. Changing the measure of probabilities, IS makes an estimator to be efficient especially for complicated model. Therefore, we consider IS for estimation of tail probability of skew normal copula model. This paper consists of two parts. First, we apply exponential twist, a usual and better IS technique, to conditional probabilities and the factors. However, this procedure does not always guarantee enough variance reduction. Such result indicates the further consideration of choosing IS factor density.

Faced with this problem, a variety of approaches has recently been proposed in the literature (Capriotti 2008, Glasserman et al 1999, Glasserman and Li 2005). The better choices of IS density can be roughly classified into two kinds of strategies. The first strategy depends on choosing optimal shift. The optimal drift is decided by using different approximation methods. Such strategy is shown in Glasserman et al 1999, or Glasserman and Li 2005.

The second strategy, as shown in Capriotti (2008), considers a family of factor probability densities which depend on a set of real parameters. By formulating in terms of a nonlinear optimization problem, IS density which is not limited the determination of drift is then determinate. The method that searches for the optimal parameters, however, incurs another efficiency problem. To keep the method efficient, particular care for robust parameters estimation needs to be taken in preliminary Monte Carlo simulation. This leads method to be more complicated.

In this paper, we describe an alternative strategy that is straightforward and flexible enough to be applied in Monte Carlo setting. Indeed, our algorithm is not limited to the determination of optimal drift in Gaussian copula model, nor estimation of parameters of factor density. To exploit the similar concept developed for basket default swap valuation in Chiang, Yueh, and Hsie (2007), we provide a reasonable guess of the optimal sampling density and then establish a way different from stochastic approximation to speed up simulation.

Finally, we provide theoretical support for single factor model and take this approach a step further to multifactor case. So we have a rough but fast approximation that execute entirely with Monte Carlo in general situation. We support our approach by some portfolio examples. Numerical results show that such algorithm is more efficient than general Monte Carlo simulation.

Keywords: Monte Carlo simulation; Importance Sampling; Portfolio credit risk; Variance reduction.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

The impressive development of the securities markets has led financial institutions to quantify their risk by stochastic models. A main component of financial risk is credit risk, in particular for rare event aspect and complex dependence between defaults of multiple obligors. By referring to losses resulting from the default of obligor, the bank and other institutions therefore make a contractual payment for structural financial products.

An important feature of modern credit risk management is to capture the effect of dependence among obligors. In the development of commercial models, the dependence structure among obligors is specified through a set of “systematic factors”. It is so-called factor copula model approach, which is originally associated with J. P. Morgan’s CreditMetrics system. To match the observed financial data, the methodology to find an adequate factor copula to model dependencies becomes very popular. There are a lot of papers to address better empirical fits of observed data by copula factor models. Examples are the normal copula model in Gupta, Finger& Bhatia (1997), the Student-t copula in Schloegl and O’Kane (2005), the double t distribution copula in Hull and White (2004) and Marshall-Olkin copula in Andersen and Sidenius (2005).

In practice, normal and Student-t copula models are two of the most widely used models. It has been incorporated into many popular risk management systems. However, more empirical works have argued the leptokurtic and asymmetric factor distribution rather than symmetric distribution. In fact, the feature of leptokurtic and asymmetry leads to significant inaccuracies in assessing the probability of extreme cases like large portfolio losses and default threshold of obligor. This result shows the importance of choosing an appropriate factor distribution for application of copula model.

In our opinion, the skew normal (SN) distribution can come up with the leptokurtic and asymmetric features. However, like most approximations in different copula cases (Glasserman (2004); Kostadinov (2005)), there are no closed form analytical results which provide the error bound of estimation. So we need a viable alternative which not only assess the performance of default but provide more information of error.

Monte Carlo simulation is the most widely used in the estimation of default. It has the advantage of being very general and disadvantage of being slowly. For estimating rare event default probability, the generalization of this method can serve the complicated copula well, but disadvantage cause time-consuming problem. This motivates research on variance reduction methods like IS to increase simulation efficiency.

For normal copula, a special case of skew normal factor model, Glasserman and Li (2005) (henceforth GL) propose a process by applying two step IS. Such approach speed up the occurrence of default event to achieve optimality of simulation. However, portfolio with SN factor is unlike the normal case, the asymmetry of SN distribution leads the procedure developed in GL is not applicable here. Surprisingly, the application of exponential twisted shift, which is a usual and better IS technique, does not always guarantee efficient variance reduction. Hence, we need to search a new way to devise IS algorithm.

Different from only considering adjustment of parameters, our approach emphasize on choosing “form” of IS density distribution. Our procedure consists of two parts. For the first step, we exploit independence property to apply IS technique to conditional probability. By this, we can reduce the part of total variance. For the second step, we eliminate the linear part of variability resulting from first step and simultaneously minimize residual volatility. Combine the two step, we then have an efficient algorithm.

The rest of this article is organized as follows. In the next chapter, In addition to the introduction of the credit risk copula model and importance sampling method, we also present the brief properties of skew normal distribution. In Chapter 3, we review two-step IS method in GL and modify the procedure of applying exponential twist technique to factor. The modified procedure does not remain well behaved in different shape parameter setting. Next, we extend a CYH procedure and build an efficient algorithm in Chapter 4. Numerical examples are illustrated in Chapter 5 and finally the concluding remarks are presented in Chapter 6.

Chapter 2 Portfolio Credit Risk Models

Credit portfolio models can be divided into reduced-form models and structural models. Discussions of several models have been put in the literatures of Crouhy et al.(2000), Bluhm et al.(2002) and McNeil et al.(2005). Different credit risk models differ in the mechanisms they use to capture dependence among obligors. In this paper, we consider the model approach similar to CreditMetrics™ (as in Gupta et al (1997) and Li (2000)), which is based on foundational work of Merton(1974). The default setting is incurred when the obligor's shortage exceeds a default threshold.

2.1 The Portfolio Loss Distribution

Consider a portfolio with m obligors, for the ith obligor, c_i and X_i denote the exposure and status respectively. The exposure c_i may be assumed to be stochastic. For sake of simplicity, we will assume c_i to be deterministic and refer the reader to Glasserman, Kang, Shahabuddin (2008) for stochastic case. The ith obligor default if X_i exceed the threshold x_i, then the portfolio loss L_m is

$$L_m = \sum_{i=1}^{m} c_i I\{X_i > x_i\} \tag{2.1}$$

In practice, threshold x_i of ith obligor is chosen according to the marginal default probability p_i so that $P(X_i > x_i) = p_i$. This value p_i is usually set based on the average historical default frequency with similar credit profiles. In the credit risk context, X_i is usually given a financial interpretation. By such framework, L_m

models the loss of a portfolio of m obligors.

Our interest is in measuring the tail behavior of L_m, particularly for rare event. Since it is impossible to exactly compute the probability of large portfolio losses, we consider asymptotic regime which supports an analysis. Such regime is relevant to portfolios of high rated obligor, or measuring risk over a short period. We assume the default threshold for the individual obligor is $x_i = b_i \sqrt{m}$, where marginal default of each obligor decrease when the number of obligors m increase. Our goal is to estimate the probability of $\{L_m > mq\}$, particular at large value of m and q.

In the framework of copula, the dependence between defaults is determined by the correlation structure between X_j and X_k, $j \neq k$. In general, the dependence structure is specified through a factor form. For example, if we set $X_k = \rho f + \varepsilon_k$ for $\rho \in \Re$ where f and ε_k are normal random variable, we obtain the dependence structure introduced in CreditMetrics™. In addition, let $X_k = \varepsilon_k / (\rho f)$ with $\rho > 0$ and f and ε_k follow Gamma and exponential distributions respectively, we get a way of introducing alternative dependence structure shown in Credit Suisse.

Here, we consider the dependence structure whose correlation is determined through a linear form, that is

$$X_i = \sum_{j=1}^{d} a_{ij} Z_j + \sqrt{1 - \sum_{j=1}^{d} a_{ij}^2}\, \varepsilon_i \qquad , i = 1, \cdots, m$$

where $Z_1,\cdots,Z_d$ and ε_i are independent random variables. The systematic factor $Z_1,\cdots,Z_d$ affects multiple obligors simultaneously, but idiosyncratic factor ε_i only influences the ith obligor. The coefficient a_{ij} is the loading for the jth factor and $\sum_{j=1}^{d} a_{ij}^2 < 1$. Those loadings are assumed to be nonnegative. Although the condition is not essential, the limitation can simplify our discussion, especially for large losses occur primarily because of highly positively correlated structure.

In the following section, we will briefly introduce properties of distribution in SN copula model.

2.2 Skew Normal Distribution and Its Properties

Let ϕ and Φ to be the standard normal probability density function and cumulative density function respectively. The density function of a random variable Z_j is given by

$$f(z \mid \mu;\sigma;\lambda) = \frac{2}{\sigma}\phi(\frac{z-\mu}{\sigma})\Phi(\lambda\frac{z-\mu}{\sigma})$$

, which is called SN distribution with location parameter $\mu \in \Re$, scale parameter $\sigma > 0$, and shape parameter $\lambda \in \Re$, denoted by $Z \sim SN(\mu,\sigma^2,\lambda)$. Let $\lambda = 0$, we obtain the normal density. As $\lambda \to \infty$, it converges pointwise to half-normal density. The SN distribution was first introduced by O'Hagan and Leonard (1976) as a prior distribution for estimating a normal location parameter. It can be applied to different fields such as economics, psychometry and so on. The moment generating function of $SN(\mu,\sigma^2,\lambda)$ is given by

$$M(t\mid\mu;\sigma;\lambda)=2\exp(\mu t+\frac{(\sigma t)^2}{2})\Phi(\frac{\lambda}{\sqrt{1+\lambda^2}}\sigma t)$$

$SN(0,1,\lambda)$ is called the standard skew normal distribution. Some main properties of SN distribution are

1. if $Z\sim SN(0,1,\lambda)$, then $-Z\sim SN(0,1,-\lambda)$.
2. if $Z\sim SN(0,1,\lambda)$, then $Z^2\sim\chi_1^2$.
3. $E(Z)=\mu+\sigma(\sqrt{2/\pi})(\lambda/\sqrt{1+\lambda^2})$.
4. $Var(Z)=\sigma^2\{1-(2/\pi)\lambda^2/(1+\lambda^2)\}$.
5. if $Z_1\sim SN(0,1,\lambda)$ and $Z_2\sim SN(0,1,0)$, then

$$\rho Z_1+\sqrt{1-\rho^2}Z_2\sim SN(0,1,\frac{\rho\lambda}{\sqrt{1+\lambda^2(1-\rho^2)}})$$

For more properties of the skew normal distribution, we refer the reader to Azzalini (1985); Gupta, Nguyen and Sanqui (2004); or Arnold and Lin(2004).

Consider $Z_j\sim SN(\mu_j,\sigma_j^2,\lambda_j)$, $j=1,\cdots,d$ and $\varepsilon_i\sim SN(0,1,0)$, we can generate Z_j, $j=1,\cdots,d$ and ε_i in each replication for straightforward simulation. We compute the value X_i and determine whether the i th obligor default. From the default setting of every obligor, we get the portfolio loss (2.1) and then evaluate the probability of $\{L_m > mq\}$. However, Monte Carlo estimator can't achieve fixed relative precision when the value of threshold becomes large. This makes variance reduction methodologies potentially attractive. In the next chapter, we will introduce one method to make Monte Carlo effective.

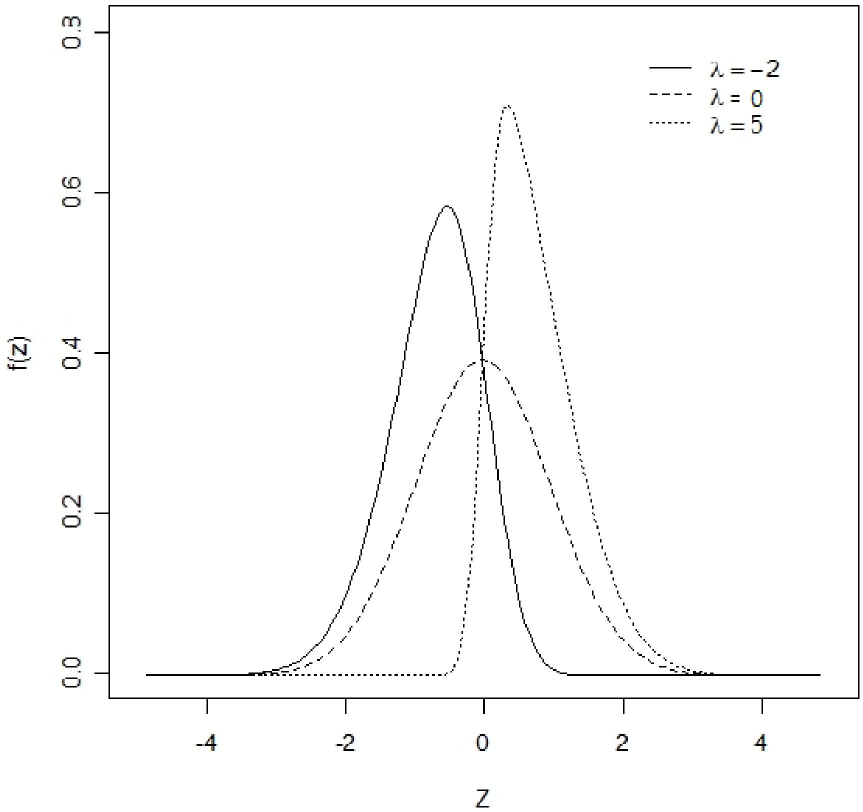

Fig. 2-1 *The density function of* $SN(0,1,\lambda)$

Chapter 3 Variance Reduction Methodology

Although generally easy to be implemented, Monte Carlo simulations are infamous for being slow. Stochastic outcome is always affected by a statistical error which can generally be reduced to the expected accuracy by iterating the procedure for a long enough time. Usually, in order to reduce the error by a factor of ten one has to spend one hundred times as much computer time. This result contradicts practical necessity.

Several approaches to accelerate the efficiency of simulation, such as control variates, antithetic variables, and IS, have been proposed over the years. The goal of these technique is to reduce the variance of replications so that an expected level accuracy can be obtained with a smaller number samples. Antithetic variables and control variates are the most commonly used variance reduction techniques, but their effectiveness varies largely across applications, and is sometimes rather limited.

Different antithetic variables and control variates, IS generally involves a bigger implementation effort and is less straightforward to include in general Monte Carlo framework. Furthermore, when used improperly, such method will "increase" the variance of estimator. That is why it has not been employed much in professional contexts until recently. For all that, its powerful variance reduction is potentially attractive.

3.1 IS Method

IS method is a standard approach of variance reduction in Monte Carlo methods. The idea behind IS method is to reduce the statistical uncertainty of a Monte Carlo calculation by focusing on the important region of space from which the random samples are drawn. For example, suppose we have to evaluate $E_f[G(X)]$ where $G(\cdot)$ is a positive measurable function with respect to the probability space and f is the density function of X, we hence have another representation

$$E[G(X)] = \int G(x) f(x) dx$$

$$= \int G(x) \frac{f(x)}{f^*(x)} f^*(x) dx$$

where f^* is another density function of X. The ratio is called likelihood ratio or Randon-Nikodym derivative. One can sample X from new density and obtain the unbiased estimator $G(x)\frac{f(x)}{f^*(x)}$. Its variance, then, is shown to be

$$E\left[G^2(X)\left(\frac{f(X)}{f^*(X)}\right)^2\right] - E^2[G(X)]$$

Indeed, we have the following optimal IS density function f_{opt} to achieve zero variance

$$f_{opt}(x) = \frac{1}{E[G(X)]} G(x) f(x)$$

Such a density exists, but it is not feasible to be found unless the desired quantity is known from the outset. Much of the literature on IS technique are focused on methods of choosing a reasonable approximation of zero-variance IS density (Sadowsky and Bucklew (1990); Glasserman et al (1999); Capriotti (2008)). The effectiveness of IS mainly depend on how close between new density and zero-variance density is. Different methods of approximation to zero-variance density will lead to various computational performances. In view of practicability consideration, we turn our attention to the apparently weaker notations of effectiveness.

Consider an estimation of $E[G(X_m)]$ where $G(X_m)$ is a function which decrease to zero as $m \to \infty$. Then an estimator $G(X_m)\frac{f(X_m)}{f^*(X_m)}$ is said to be of *bounded relative error* if it satisfies the requirement

$$\limsup_{m\to\infty} \frac{\sqrt{Var[G(X_m)\frac{f(X_m)}{f^*(X_m)}]}}{E[G(X_m)]} < \infty$$

Additionally, an estimator is called *logarithm asymptotically efficient* or *asymptotically optimal* if it satisfies the requirement

$$\lim_{m\to\infty} \frac{\ln E\left[G^2(X_m)\left(\frac{f(X_m)}{f^*(X_m)}\right)^2\right]}{\ln E[G(X_m)]} = 2$$

An estimator with bounded relative error can remain the number of replication bounded in a fixed bounded relative error. However, asymptotically optimal only ensure that the rate of decay of second moment achieves twice that of itself. By Jensen's inequality, this is the fastest possible rate of decrease for any unbiased estimator. By the simple algebra, we know that an estimator with bounded relative error is also asymptotically optimal. General Monte Carlo can't achieve asymptotically optimal. For example, let $p_m = P(X > m)$ where $X \sim N(0,1)$, then a general Monte Carlo estimator $I\{X > m\}$ has the following result

$$\lim_{m\to\infty} \frac{\ln E[I^2\{X > m\}]}{\ln E[I\{X > m\}]} = 1$$

3.2 IS Conditional on SN Factor

In this section, we apply the one-step of GL to the credit portfolio with skew normal factors. To keep the notation simple, we restrict our attention to single factor homogeneous model, that is $c_i = d = 1$, $\rho_i = \rho$, $b_i = b$ and X_i is given by

$$X_i = \rho Z + \sqrt{1-\rho^2}\,\varepsilon_i$$

Thus the total loss L_m can be written as

$$L_m = \sum_{i=1}^{m} I\{\rho Z + \sqrt{1-\rho^2}\,\varepsilon_i > b\sqrt{m}\}$$

Let $Y_i = I\{\rho Z + \sqrt{1-\rho^2}\,\varepsilon_i > b\sqrt{m}\}$. Conditioning on $Z = z$, Y_i is a Bermoulli random variable and the conditional default probability $p(z)$ is given by

$$p(z) = P(Y_i = 1 \mid Z = z)$$
$$= \Phi(\frac{\rho z - b\sqrt{m}}{\sqrt{1-\rho^2}})$$

The joint probability of $(Y_1, \cdots, Y_m)$ is then given by

$$\prod_{i=1}^{m} (p(z))^{Y_i} (1 - p(z))^{1-Y_i}$$

Consider a new probability $q(z;\theta(z))$ which is specified by

$$p_{\theta(z)}(z) = \frac{p(z)e^{\theta(z)}}{\exp(\psi(\theta(z);z))} \tag{3.1}$$

where $\psi(\theta(z);z) = \ln(1 + p(z)(e^{\theta(z)} - 1))$. If we replace each probability $p(z)$ with a new probability $p_{\theta(z)}(z)$, then the estimation of $E[I\{L_m > mq\}]$ can be written as

$$E\left[I\{L_m > mq\} \prod_{i=1}^{m} (\frac{p(Z)}{p_{\theta(Z)}(Z)})^{Y_i} (\frac{1-p(Z)}{1-p_{\theta(Z)}(Z)})^{1-Y_i} \right]$$

(3.1) is called *exponential twist*. If $\theta(z) > 0$, then $p_{\theta(z)}(z) > p(z)$; the original probability correspond to $\theta(z) = 0$. The new measure can increase the default probability and so decrease the variance resulting from stochastic volatility.

Let $\psi_{L_m}(\theta(z), z) = \sum_{i=1}^{m} \psi(\theta(z); z)$, the corresponding likelihood ratio is simplified into

$$\prod_{i=1}^{m}(\frac{p(z)}{p_{\theta(z)}(z)})^{Y_i}(\frac{1-p(z)}{1-p_{\theta(z)}(z)})^{1-Y_i}=e^{-\theta(z)L_m+\psi_{L_m}(\theta(z),z)}$$

, then we have an unbiased estimator

$$I\{L_m>mq\}e^{-\theta(z)L_m+\psi_{L_m}(\theta(z),z)} \tag{3.2}$$

for $P(L_m>mq)$. It remains to choose $\theta(z)$ to reduce the variance of (3.2). We know the fact that a key element of variance reduction is based on minimizing the second moment. It is difficult to solve this problem directly, but minimizing the upper bound of the second moment is easy. For $\theta(z)$, we know

$$E[I\{L_m>mq\}e^{-2\theta(z)L_m+2\psi_{L_m}(\theta(z),z)}\mid Z=z]\le e^{-2\{\theta(z)mq-\psi_{L_m}(\theta(z),z)\}}$$

Note that the function $\psi_{L_m}(\theta(z),z)$ is strictly convex in $\theta(z)$ and pass through the origin. So the function $\theta(z)mq-\psi_{L_m}(\theta(z),z)$ is a concave function and the supremun is attained at just one point, which we denote by $\theta_m(z;q)$. If $p(z)\ge q$, then the maximum value occurs at $\theta_m(z;q)=0$; otherwise, it occurs at the unique solution of

$$\frac{\partial}{\partial\theta(z)}\psi_{L_m}(\theta_m(z;q),z)=mq$$

In this discussion, $\theta_m(z;q)$ may be viewed as a measure of the conditional rarity of the set $\{L_m > mq\}$. If $\{L_m > mq\}$ is not rare event, we generate the $Y_i \mid Z = z$ from the original probability; otherwise we twist by $\theta_m(z;q)$. Observe the set $\{L_m > mq\}$, any element of $\{L_m > mq\}$ has the property

$$\theta_m(z;q)L_m \geq \theta_m(z;q)mq,$$

,then we have the lower bound

$$-\theta_m(z;q)L_m + \psi_{L_m}(\theta_m(z;q),z) \leq F_m(z)$$

Here $F_m(z) = -\theta_m(z;q)mq + \psi_{L_m}(\theta_m(z;q),z) \leq 0$. This result shows that the asymptotic of upper bound depend on the point q when we apply IS to conditional probability. This unique point q is called the *dominating point*. Essentially, the decreasing rate of $F_m(z)$ determines whether the new estimator achieve asymptotically optimal. Observe that the equation

$$\begin{aligned} E[L_m \mid Z = z] &= \frac{\partial}{\partial\theta(z)}\psi_{L_m}(\theta_m(z;q),z) \\ &= mq \end{aligned}$$

holds. This fact indicates that the mean value of $\frac{L_m}{m} \mid Z = z$ will be shifted to the dominating point q if $\{L_m > mq\}$ becomes rare event. More details about dominating point we refer the readers to Ney (1983) or Sadowsky and Bucklew (1990).

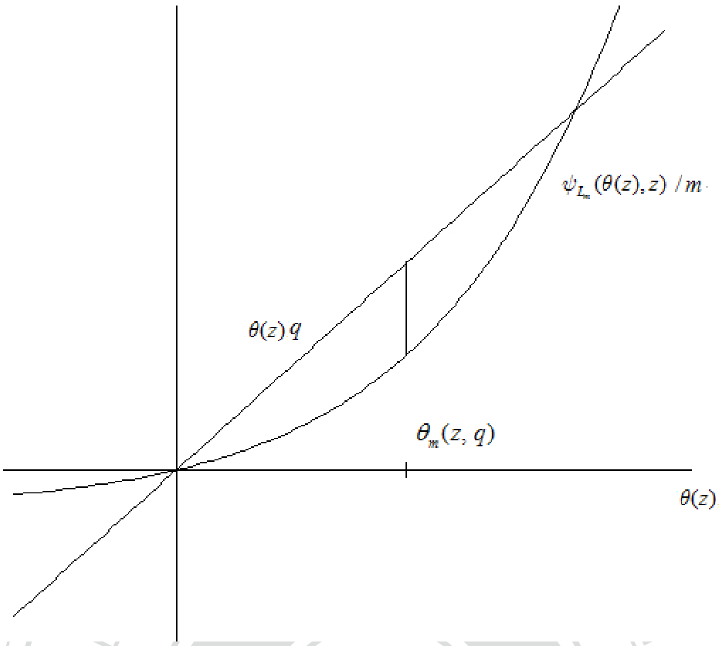

Fig. 3-1 *A specific example with* $\rho=0.3$, $b=0.2$, $m=100$, $q=0.1$ *and* $z=0.5$. *The tangent to* $\psi_{L_m}(\cdot,z)/m$ *at the point* $\theta_m(z;q)$ *is* q. *The vertical distance from* $\psi_{L_m}(\theta_m(z;q),z)/m$ *to the line through the original is* $-F_m(z)/m$.

Note that the value of $\theta_m(z;q)$ is chosen to reduce the variance of $P(L_m > mq \mid Z)$ rather than $P(L_m > mq)$. In GL, failure of asymptotical optimal results from the extra variability of Z. To analyze further, we should first realize how $F_m(z)$ behave. Set loading ρ to be 0.1, 0.3, and 0.8 respectively, $-F_m(z)/m$ is shown in Fig. 3-2.

In Fig. 3-2, it is obvious the larger value ρ is, the larger influence Z has. One step IS turns out to be less effective if the structure of $F_m(z)$ has a width which depart from a constant. The behavior of $F_m(z)$ indicates irrationality to neglect the effect from the factor. Hence, we need to exploit proper IS method again to reduce the impact of factor.

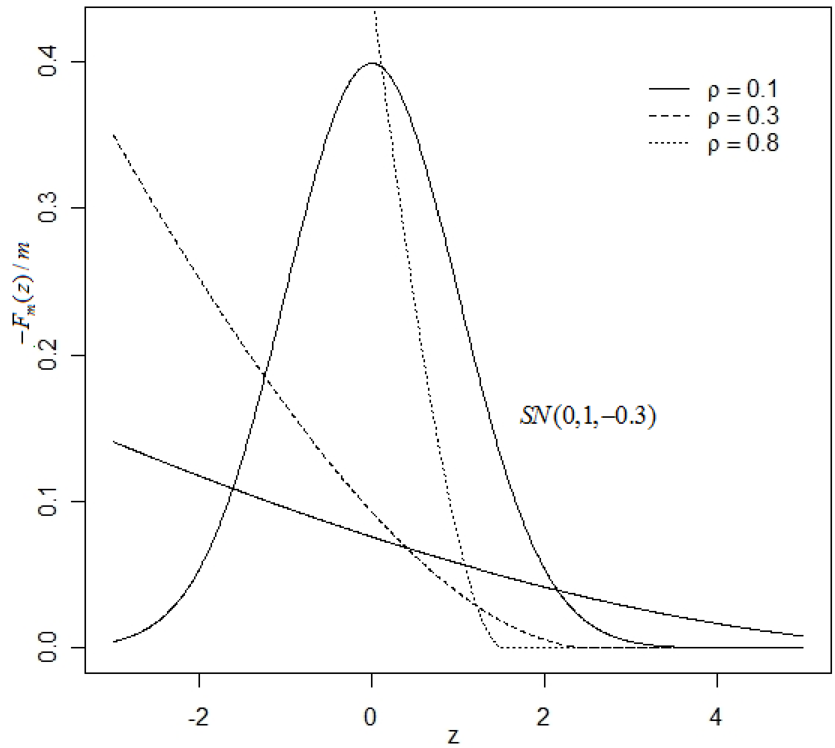

Fig. 3-2 *Comparisom of* $-F_m(z)/m$ *under different loading values with* $b=0.2$, $m=100$, $q=0.1$.

3.3 IS For SN Factor

As discussion in section 3.2, the key to improve efficiency of (3.2) is based on the elimination of residual randomness. Consider the (3.2) with $\theta(z)=\theta_m(z;q)$, any estimator $\widehat{p}_{mq}$ has the variance decomposition as following

$$Var[\widehat{p}_{mq}]=E[Var[\widehat{p}_{mq}\,|\,Z]]+Var[E[\widehat{p}_{mq}\,|\,Z]]$$

Applying IS to conditioning Z only makes $E[Var[\widehat{p}_{mq} \mid Z]]$ small. To get further improvement of efficiency, GL focus on the second term in the variance decomposition. By simple algebra, we know that the zero-variance IS density for $Var[E[\widehat{p}_{mq} \mid Z]]$ is

$$\frac{1}{E[I\{L_m > mq\}]} E[I\{L_m > mq\} \mid Z = z] f(z \mid \mu; \sigma; \lambda)$$

It is pity that sampling from this density is generally infeasible because of the normalization constant. For $(\mu, \sigma, \lambda) = (0,1,0)$, GL thus suggest using original distribution with a appropriate mode as optimal density. Rather than choose IS density arbitrarily, it is intuitively clear that shifting the mode makes the likelihood ratio inside expectation to be small. For symmetric density, such strategy may make a substantial variance reduction. Whenever zero variance density cannot be approximated only by shifting the mode, however, this algorithm becomes less beneficial. For instance, when the structure of $E[I\{L_m > mq\} \mid Z = z] f(z \mid \mu; \sigma; \lambda)$ has a width which is very different from original density, shifting a drift will turn out to be ineffective mechanism.

Faced with a similar problem, Capriotti (2008) uses Levenberg-Marquardt method to provide a reasonable IS density which is not limited the determination of drift. The implementation to determine the optimal parameters, however, incurs another efficiency problem. To keep the algorithm efficient, particular care for robust parameter estimation needs to be taken in preliminary Monte Carlo simulation. This leads algorithm to be more complicated.

Instead of solving robust problem, we adopt another strategy to vanish randomness from Z. Let a new estimator applying IS for factor and conditional on factor to be

$$I\{L_m > mq\}e^{-2\{\theta_m(Z;q)L_m - \psi_{L_m}(\theta_m(Z;q),Z)\}}W(Z) \tag{3.3}$$

Here $W(Z) = f(z\mid\mu;\sigma;\lambda)/f^*(z\mid\mu;\sigma;\lambda)$ and $f^*(z\mid\mu;\sigma;\lambda)$ denotes the IS density. We put emphasis on the variance of (3.3) rather than variance decomposition. Observe the second moment of (3.3)

$$\begin{aligned}
&E\left[I\{L_m > mq\}e^{-2\theta_m(z;q)L_m + 2\psi_{L_m}(\theta_m(z;q),z)}W^2(Z)\right] \\
&\qquad \le E\left[I\{L_m > mq\}e^{2F_m(Z)}W^2(Z)\right] \\
&\qquad \le \int e^{2F_m(z)}W^2(Z)f^*(z\mid\mu;\sigma;\lambda)dz
\end{aligned}$$

Directly minimizing second moment is difficult, a surrogate to guide proper IS density factor is required and thus we consider the upper bound of second moment. To avoid integrating intricate function $F_m(z)$, we choose a different approximation. By the simple differentiation, we know $F_m(z)$ is concave and then we have a loose upper bound by the first order Taylor expansion at point t_m.

$$\begin{aligned}
\int e^{2F_m(z)}W^2(Z)f^*(z\mid\mu;\sigma;\lambda)dz &\le \int e^{2\{F_m(t_m)+F_m'(t_m)(z-t_m)\}}W^2(Z)f^*(z\mid\mu;\sigma;\lambda)dz \\
&= E[\{e^{\{F_m(t_m)+F_m'(t_m)(Z-t_m)\}}W(Z)\}^2]
\end{aligned}$$

Considering Jensen's inequality, the inequality holds if

$$e^{F_m(t_m)+F_m^{'}(t_m)(Z-t_m)}W(Z) = E[e^{F_m(t_m)+F_m^{'}(t_m)(Z-t_m)}W(Z)]$$

$$= E[e^{F_m(t_m)+F_m^{'}(t_m)(Z-t_m)}],$$

then the formulation yields

$$f^*(z\,|\,\mu;\sigma;\lambda) = \frac{e^{F_m^{'}(t_m)z}}{M_Z(F_m^{'}(t_m))} f(z\,|\,\mu;\sigma;\lambda) \tag{3.4}$$

where $M_Z(\cdot)$ denotes the moment generating function of Z. If we consider the exponential twist density of Z as

$$f^*(z\,|\,\mu;\sigma;\lambda) = \frac{e^{t_m z}}{M_Z(t_m)} f(z\,|\,\mu;\sigma;\lambda), \tag{3.5}$$

this connection of (3.4) and (3.5) implies that we can design a new exponential twisted IS density where $t_m = F_m^{'}(t_m)$, namely $t_m = \arg\max_t \{F_m(t) - t^2/2\}$

Once we have selected the new IS density of Z, the algorithm proceeds as follows:

1. Compute $t_m = \arg\max_t \{F_m(t) - t^2/2\}$

2. Sample Z from $f^*(z \mid \mu;\sigma;\lambda)$

3. Twisting the conditional default probability

4. Return the estimator

$$\hat{p}_{ET} = I\{L_m > mq\} e^{-\theta_m(Z;q)L_m + \psi_{L_m}(\theta_m(Z;q),Z)} e^{-t_m Z + \log M_Z(t_m)}$$

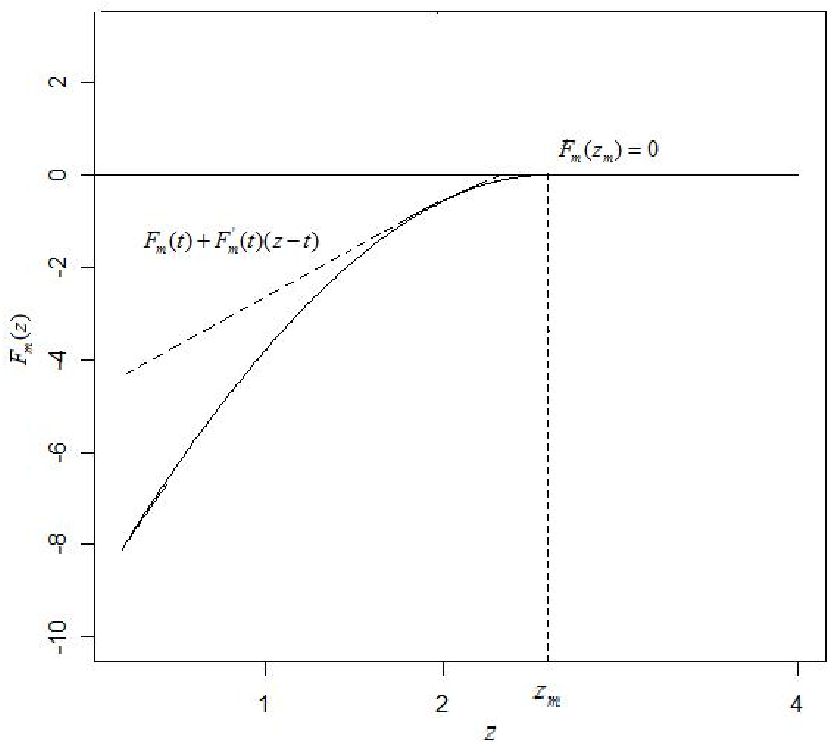

Fig. 3-3 *Graph of* $F_m(t) + F_m^{'}(t)(z-t)$ *for a single factor with* $b = 0.2$, $\rho = 0.3$, $q = 0.1$ *and* $m = 1{,}000$.

By simple algebra, we know that the new IS distribution for Z is the closed skew normal $CSN(t_m, 1, \lambda, -\lambda t_m, 1)$ if the original distribution is $SN(0,1,\lambda)$. In fact, the appropriate IS distribution is not limited to the original distribution family and this result shows the difficulty in searching the optimal IS density. For more properties of closed skew normal one can see Graciela et al (2004).

Note that if $\lambda = 0$, the choice of IS density coincides with the result of GL , that is a normal density with mean t_m, namely $CSN(t_m, 1, 0, 0, 1)$.

Theorem 3.1

Consider a single factor homogeneous portfolio with the factor $Z \sim SN(0,1,\lambda)$ and $\varepsilon_i \sim SN(0,1,0)$. Suppose the default and loss threshold are $b\sqrt{m}$ and mq respectively. Then the estimator $\hat{p}_{ET}$ satisfy

(a)For $\lambda \geq 0$

$$\lim_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] = -\frac{b^2}{2\rho^2}$$

$$\lim_{m\to\infty} \frac{1}{m} \log E[(\hat{p}_{ET})^2] = -\frac{b^2}{\rho^2}$$

(b)For $\lambda < 0$

$$\lim_{m\to\infty} \frac{1}{m} \log E\{I\{L_m > mq\}\} = -\frac{b^2}{2\rho^2}(1+\lambda^2)$$

$$\lim_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} = -\frac{b^2}{\rho^2}(\frac{1+2\lambda^2}{1+\lambda^2})$$

Proof: The result would follow from a similar discussion in GL (2004). To start, we consider lower bound and upper bound of liminf and limsup respectively.

(a) $\lambda \geq 0$: First, we show that

$$\liminf_{m\to\infty}\frac{1}{m}\log E[I\{L_m > mq\}] \ge -\frac{b^2}{2\rho^2}$$

By conditional property, for any $\upsilon > 0$, we have the following result

$$\begin{aligned} E[I\{L_m > mq\}] &= E[E[I\{L_m > mq\} \mid p(Z) \ge q+\upsilon]] \\ &= P(L_m > mq \mid p(Z) \ge q+\upsilon)P(p(Z) \ge q+\upsilon) \\ &\ge P(L_m > mq \mid p(Z) = q+\upsilon)P(p(Z) > q+\upsilon) \end{aligned}$$

The inequality holds because L_m is binomially distributed with parameter m and $p(Z)$.

Applying the lower bound (3.62) of Johnson et al. (1993), we have a lower bound for the conditional probability $P(L_m > mq \mid p(Z) = q+\upsilon) \ge 1/2$ .Substituting the result into the lower bound, we have

$$\begin{aligned} E[I\{L_m > mq\}] &\ge \frac{1}{2}P(p(Z) > q+\upsilon) \\ &= \frac{1}{2}P(Z > \frac{b\sqrt{m}+\sqrt{1-\rho^2}\,\Phi^{-1}(q+\upsilon)}{\rho}) \\ &= \frac{1}{2}\{1-F_Z(\frac{b\sqrt{m}+\sqrt{1-\rho^2}\,\Phi^{-1}(q+\upsilon)}{\rho})\} \end{aligned}$$

where $F_Z(\cdot)$ denotes the cdf of Z and

$$\liminf_{m\to\infty}\frac{1}{m}\log E[I\{L_m > mq\}] \geq \liminf_{m\to\infty}\frac{1}{m}\log\{1-F_Z(\frac{b\sqrt{m}+\sqrt{1-\rho^2}\Phi^{-1}(q+\upsilon)}{\rho})\}$$

Applying l'Hospital's rule, we obtain

$$\liminf_{m\to\infty}\frac{1}{m}\log\{1-F_Z(\frac{b\sqrt{m}+\sqrt{1-\rho^2}\Phi^{-1}(q+\upsilon)}{\rho})\}$$

$$=-\frac{b}{\rho}\liminf_{m\to\infty}\frac{\frac{b\sqrt{m}-\sqrt{1-\rho^2}\xi_{1-q}(1+\delta/\xi_{1-q})}{2\rho}+\frac{\rho}{2\sqrt{m}}}{\sqrt{m}\Phi(\lambda\frac{b\sqrt{m}-\sqrt{1-\rho^2}\xi_{1-q}(1+\delta/\xi_{1-q})}{2\rho^2})}$$

$$=-\frac{b}{\rho}\liminf_{m\to\infty}\{\frac{b\sqrt{m}-\sqrt{1-\rho^2}\xi_{1-q}(1+\delta/\xi_{1-q})}{2\rho\sqrt{m}\Phi(\lambda\frac{b\sqrt{m}-\sqrt{1-\rho^2}\xi_{1-q}(1+\delta/\xi_{1-q})}{2\rho^2})}+\frac{\rho}{2m\Phi(\lambda\frac{b\sqrt{m}-\sqrt{1-\rho^2}\xi_{1-q}(1+\delta/\xi_{1-q})}{2\rho^2})}\}$$

$$=-\frac{b^2}{2\rho^2}$$

which proves the formulation.

Next we show that

$$\limsup_{m\to\infty}\frac{1}{m}\log E\{(\hat{p}_{ET})^2\} \leq -\frac{a^2}{\rho^2}$$

Write $E\{(\hat{p}_{ET})^2\}$ as

$$E\{(\hat{p}_{ET})^2\} = E\{I\{L_m > mq\}\exp\{-2\theta_m(Z;q)L_m + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + 2\ln M_Z(t_m)\}\}$$

$$\leq 4E[\exp(-2\theta_m(Z;q)mq + 2\psi_{L_m}(\theta_m(Z;q),Z)) - 2t_m Z + t_m^2\}\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)]$$

$$\leq 4E\{\exp\{2F_m(Z) - 2t_m Z + t_m^2\}\}\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)$$

where $F_m(z) = -\theta_m(z;q)mq + \psi_{L_m}(\theta_m(z;q),z)$.

By differentiation, we know that $F_m(\cdot)$ is an increasing and concave function, thus we know for any $t_m \in \Re$

$$F_m(z) \leq F_m(t_m) + F_m'(t_m)(z - t_m)$$

and

$$E\{(\hat{p}_{ET})^2\} \leq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E\{\exp\{2(F_m(t_m) + F_m'(t_m)(Z - t_m)) - 2t_m Z + t_m^2\}\}$$

$$\leq 4\exp\{2(F_m(t_m) - \frac{t_m^2}{2})\}$$

The second inequality holds because $F_m'(t_m) = t_m$ and $\Phi(\cdot) \leq 1$. Consider

$$z_m = \frac{b\sqrt{m} + \sqrt{1-\rho^2}\Phi^{-1}(q)}{\rho}$$

Thus we have $p(z_m) = q$ and $p(z_m) \geq q$ for any $z \geq z_m$. Next, we show that for small $\varsigma > 0$, we can find m_1 that $t_m \in (z_m(1-\varsigma), z_m)$ if $m > m_1$. It suffices to show

$$F_m^{'}(z_m(1-\varsigma)) - z_m(1-\varsigma) > 0 \quad \text{and} \quad F_m^{'}(z_m) - z_m < 0$$

We know that the second inequality holds because $p(z_m) = q$ and

$$F_m^{'}(z) = m(\frac{q - p(z)}{p(z)(1-p(z))})\phi(\frac{\rho z - b\sqrt{m}}{\sqrt{1-\rho^2}})\frac{\rho}{\sqrt{1-\rho^2}}$$

where $\phi(\cdot)$ denotes the density function of standard normal random variable. For the first inequality, we get

$$F_m^{'}(z_m(1-\varsigma)) = m(\frac{q - p(z_m(1-\varsigma))}{p(z_m(1-\varsigma))(1-p(z_m(1-\varsigma)))})\phi(\frac{\rho z_m(1-\varsigma) - b\sqrt{m}}{\sqrt{1-\rho^2}})\frac{\rho}{\sqrt{1-\rho^2}}$$

By l'Hospital's rule, we thus have

$$\frac{q - p(z_m(1-\varsigma))}{p(z_m(1-\varsigma))(1-p(z_m(1-\varsigma)))} = O(\frac{1}{\Phi(-\frac{b\sqrt{m} - \rho z_m(1-\varsigma)}{\sqrt{1-\rho^2}})})$$

Applying the property that $\phi(x)/\Phi(-x) \sim x$ if $x \to \infty$, thus we conclude that

$$F_m^{'}(z_m(1-\varsigma)) = O(m^{3/2})$$

Since $z_m = O(m^{1/2})$, we obtain the first inequality when m is large enough.

Substituting the result into the upper bound of $E\{(\hat{p}_{ET})^2\}$, we then obtain

$$E\{(\hat{p}_{ET})^2\} \le 4\exp\{2(F_m(t_m) - \frac{t_m^2}{2})\}$$

$$\le 4\exp\{2(F_m(z_m) - \frac{(z_m(1-\varsigma))^2}{2})\}$$

The second inequality holds because $F_m(z)$ is increasing function. Due to $F_m(z_m) = 0$.

we then get

$$\limsup_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} \le \limsup_{m\to\infty} \frac{-1}{m}(z_m(1-\varsigma))^2$$

$$\le \limsup_{m\to\infty} \frac{-1}{m}\{\frac{b^2 m}{\rho^2} + o(m)\}$$

$$= -\frac{b^2}{\rho^2}$$

By Jensen's inequality, we complete the proof.

(b) $\lambda < 0$:

First, we show the lower bound

$$\liminf_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] \geq -\frac{b^2}{2\rho^2}(1+\lambda^2)$$

Define

$$z_m(\upsilon) = \frac{b\sqrt{m} + \sqrt{1-\rho^2}\Phi^{-1}(q+\upsilon)}{\rho}$$

By the lower bound (3.62) of Johnson et al. (1993), we have

$$\begin{aligned} E[I\{L_m > mq\}] &= P(L_m > mq \mid p(Z) \geq q+\upsilon)P(p(Z) \geq q+\upsilon) \\ &\geq P(L_m > mq \mid p(Z) = q+\upsilon)P(p(Z) > q+\upsilon) \\ &= \frac{1}{2}P(\Phi(\frac{\rho Z - b\sqrt{m}}{\sqrt{1-\rho^2}}) \geq q+\upsilon) \\ &\geq \frac{1}{2}P(z_m(\upsilon) \leq Z \leq z_m(\upsilon)+\kappa_0) \end{aligned}$$

for any $\upsilon, \kappa_0 > 0$. Note that $z_m(\upsilon) > 0$ for m sufficiently large. Hence, the probability is lower bounded by

$$\kappa_0 \phi(z_m(\upsilon)+\kappa_0)\Phi(\lambda(z_m(\upsilon)+\kappa_0))$$

and we obtain

$$\liminf_{m\to\infty}\frac{1}{m}\log E[I\{L_m > mq\}]$$

$$\geq \liminf_{m\to\infty}\frac{1}{m}\log\phi(z_m(\upsilon)+\kappa_0)+\liminf_{m\to\infty}\frac{1}{m}\log\Phi(\lambda(z_m(\upsilon)+\kappa_0))$$

Note that

$$\liminf_{m\to\infty}\frac{1}{m}\log\phi(z_m(\upsilon)+\kappa_0)=\liminf_{m\to\infty}\frac{1}{m}\{\frac{-1}{2}(z_m(\upsilon)+\kappa_0)^2\}$$

$$=\liminf_{m\to\infty}\frac{1}{m}\{\frac{-b^2}{2\rho^2}(m+o(m))\}$$

$$=-\frac{b^2}{2\rho^2}$$

Applying l'Hospital's rule and the property that $\phi(x)/\Phi(-x)\sim x$ as $x\to\infty$, we have

$$\liminf_{m\to\infty}\frac{1}{m}\log\Phi(\lambda(z_m(\upsilon)+\kappa_0))=\liminf_{m\to\infty}\frac{1}{m}\log\Phi(-|\lambda|(z_m(\upsilon)+\kappa_0))$$

$$=\liminf_{m\to\infty}\frac{\phi(-|\lambda|(z_m(\upsilon)+\kappa_0))}{\Phi(-|\lambda|(z_m(\upsilon)+\kappa_0))}\frac{-b|\lambda|}{2\rho\sqrt{m}}$$

$$=\liminf_{m\to\infty}\frac{\phi(|\lambda|(z_m(\upsilon)+\kappa_0))}{\Phi(-|\lambda|(z_m(\upsilon)+\kappa_0))}\frac{-b|\lambda|}{2\rho\sqrt{m}}$$

$$=\liminf_{m\to\infty}\frac{-b|\lambda|}{2\rho}\{|\lambda|(z_m(\upsilon)+\kappa_0)+o(\sqrt{m})\}\frac{1}{\sqrt{m}}$$

$$=\frac{-b^2\lambda^2}{2\rho^2}$$

Combining all results, we get the formulation.

Next we show the upper bound

$$\limsup_{m\to\infty}\frac{1}{m}\log E[I\{L_m > mq\}] \le -\frac{b^2}{2\rho^2}(1+\lambda^2)$$

We know

$$\begin{aligned}
E[I\{L_m > mq\}] &\le 2E\{\Phi(\lambda Z)\exp\{-\theta_m(Z;q)mq + \psi_{L_m}(\theta_m(Z;q),Z) - t_m Z + t_m^2/2\}\} \\
&\le 2E\{\Phi(\lambda Z)\exp\{F_m(Z) - t_m Z + t_m^2/2\}\} \\
&\le 2E\{\Phi(\lambda Z)\exp\{F_m(t_m) + F_m^{'}(t_m)(Z-t_m) - t_m Z + t_m^2/2\}\} \\
&\le 2\exp\{F_m(t_m) - t_m^2/2\}E\{\Phi(\lambda Z)\} \\
&= 2\exp\{F_m(t_m) - t_m^2/2\}E\{\Phi(\lambda(Z+t_m))\}
\end{aligned}$$

For any $\zeta > 0$, if m is large sufficiently, we have

$$\begin{aligned}
E\{\Phi(\lambda(Z+t_m))\} &= E[I\{Z\ge 0\}\Phi(\lambda(Z+t_m))] + E[I\{Z<0\}\Phi(\lambda(Z+t_m))] \\
&\le \frac{1}{2}\Phi(\lambda t_m)(1-2\zeta) + \frac{1}{2}\Phi(\lambda t_m)(1+\zeta) \\
&\le \Phi(\lambda t_m)(1-\frac{\zeta}{2})
\end{aligned}$$

The inequality holds because of second mean value theorem for integral, so we know

$$E[I\{L_m > mq\}] \le 2(1-\frac{\zeta}{2})\exp\{F_m(t_m) - \frac{t_m^2}{2}\}\Phi(\lambda t_m)$$

By the similar argument of z_m and t_m, for any $\varsigma > 0$, if m is sufficiently large, we get

$$E[I\{L_m > mq\}] \le 2(1-\frac{\zeta}{2})\Phi(\lambda z_m(1-\varsigma))\exp\{-\frac{(z_m(1-\varsigma))^2}{2}\}$$

and

$$\limsup_{m\to\infty}\frac{1}{m}\log E[I\{L_m > mq\}]$$

$$\le \limsup_{m\to\infty}\frac{1}{m}\log\Phi(\lambda z_m(1-\varsigma)) + \limsup_{m\to\infty}\frac{-1}{m}\frac{(z_m(1-\varsigma))^2}{2}$$

For the second term, we know that

$$\limsup_{m\to\infty}\frac{-1}{m}\frac{(z_m(1-\varsigma))^2}{2} = \limsup_{m\to\infty}\frac{1}{m}\{\frac{-b^2 m}{2\rho^2} + o(m)\}$$

$$= \frac{-b^2}{2\rho^2}$$

Applying l'Hospital's rule and the property that $\phi(x)/\Phi(-x) \sim x$ as $x \to \infty$, we have

$$\limsup_{m\to\infty}\frac{1}{m}\log\Phi(\lambda z_m(1-\varsigma)) = \limsup_{m\to\infty}\frac{1}{m}\log\Phi(-|\lambda| z_m(1-\varsigma))$$

$$= \limsup_{m\to\infty} \frac{\phi(-|\lambda| z_m(1-\varsigma))}{\Phi(-|\lambda| z_m(1-\varsigma))} \frac{-b|\lambda|}{2\rho\sqrt{m}}$$

$$= \limsup_{m\to\infty} \frac{\phi(|\lambda| z_m(1-\varsigma))}{\Phi(-|\lambda| z_m(1-\varsigma))} \frac{-b|\lambda|}{2\rho\sqrt{m}}$$

$$= \limsup_{m\to\infty} \frac{-b|\lambda|}{2\rho} \{|\lambda| z_m(1-\varsigma) + o(\sqrt{m})\} \frac{1}{\sqrt{m}}$$

$$= \frac{-b^2\lambda^2}{2\rho^2}$$

Combining those inequalities, we obtain

$$\limsup_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] \le \frac{-b^2(1+\lambda^2)}{2\rho^2}$$

By the limsup and liminf, we have

$$\lim_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] = -\frac{b^2}{2\rho^2}(1+\lambda^2)$$

which complete the first part of proof.

Next, we show

$$\liminf_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} \ge -\frac{b^2}{\rho^2}\left(\frac{1+2\lambda^2}{1+\lambda^2}\right)$$

For any $\delta > 0$ and let $z_m(-\delta) = \dfrac{b\sqrt{m} + \sqrt{1-\rho^2}\Phi^{-1}(q-\delta)}{\rho}$, we have

$$E[(\hat{p}_{ET})^2] = 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{L_m > mq\}$$

$$\exp\{-2\theta_m(Z;q)L_m + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + t_m^2\}]$$

$$\geq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{mq < L_m < m(q+\delta)\}$$

$$\exp\{-2\theta_m(Z;q)m(q+\delta) + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + t_m^2\}]$$

$$\geq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{mq < L_m < m(q+\delta)\}I\{p(Z) \leq q\}$$

$$\exp\{-2\theta_m(Z;q)m(q+\delta) + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + t_m^2\}]$$

$$= 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{mq < L_m < m(q+\delta)\} \mid p(Z) \leq q]$$

$$E[I\{q-\delta \leq p(Z) \leq q\}\exp\{2mG_\delta(p(Z)) - 2t_m Z + t_m^2\}]$$

$$= 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{mq < L_m < m(q+\delta)\} \mid p(Z) \leq q]$$

$$E[I\{z_{m,\delta} \leq Z \leq z_m\}\exp\{2mG_\delta(p(Z)) - 2t_m Z + t_m^2\}]$$

$$\geq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{mq < L_m < m(q+\delta)\} \mid p(Z) \leq q]$$

$$\exp\{2mG_\delta(p(z_{m,\delta})) - 2t_m z_m + t_m^2\}$$

Here $G_\delta(p(z)) = -2\theta_m(Z;q)m(q+\delta) + 2\psi_{L_m}(\theta_m(Z;q),z)$ and $G_\delta(p(z))$ is increasing function of z. The equation hold because L_m and Z are independent given $p(z) \leq q$. The loss L_m has a binomial distribution with parameter m and q. Hence, by the central limit theorem, for m large enough,

$$E[I\{mq < L_m < m(q+\delta)\} \mid p(Z) \le q]$$

$$= E[I\{0 \le \frac{L_m - mq}{\sqrt{mq(1-q)}} \le \sqrt{\frac{m}{q(1-q)}}\delta\}] \ge \frac{1}{4}$$

and for all $\nu_\delta > 0$, we also have $G_\delta(p(z_{m,\delta})) \ge -\nu_\delta$. Therefore

$$\liminf_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} \ge \liminf_{m\to\infty} \frac{1}{m} \log \Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m)$$

$$+ \liminf_{m\to\infty} \frac{1}{m} \{-2m\nu_\delta - 2t_m z_m + t_m^2\}$$

Apply the result, $t_m \in (z_m(1-\varsigma), z_m)$ for any $\varsigma > 0$ if m large enough and l'Hospital's rule, by following the same steps discussed before then we get

$$\liminf_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} \ge -\frac{b^2}{\rho^2}(\frac{1+2\lambda^2}{1+\lambda^2})$$

Next, we show

$$\limsup_{m\to\infty} \frac{1}{m} \log E\{(\hat{p}_{ET})^2\} \le -\frac{b^2}{\rho^2}(\frac{1+2\lambda^2}{1+\lambda^2})$$

Consider

$$E[(\hat{p}_{ET})^2] = 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[I\{L_m > mq\}$$

$$\exp\{-2\theta_m(Z;q)L_m + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + t_m^2\}]$$

$$\leq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)\exp\{-2\theta_m(Z;q)mq + 2\psi_{L_m}(\theta_m(Z;q),Z) - 2t_m Z + t_m^2\}]$$

$$\leq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)E[\exp\{2F_m(t_m) + 2F_m^{'}(t_m)(Z - t_m) - 2t_m Z + t_m^2\}]$$

$$\leq 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)\exp\{F_m(t_m) - \frac{t_m^2}{2}\}$$

$$= 4\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)\exp\{F_m(t_m) - \frac{t_m^2}{2}\}$$

Therefore we have

$$\limsup_{m\to\infty}\frac{1}{m}\log E\{(\hat{p}_{ET})^2\} \leq \limsup_{m\to\infty}\frac{1}{m}\log\Phi^2(\frac{\lambda}{\sqrt{1+\lambda^2}}t_m)$$

$$+\limsup_{m\to\infty}\frac{1}{m}\{F_m(t_m) - \frac{t_m^2}{2}\}$$

Apply the result, $t_m \in (z_m(1-\varsigma), z_m)$ for any $\varsigma > 0$ and l'Hospital's rule, by following the same steps discussed before then we get

$$\limsup_{m\to\infty}\frac{1}{m}\log E\{(\hat{p}_{ET})^2\} \leq -\frac{b^2}{\rho^2}(\frac{1+2\lambda^2}{1+\lambda^2})$$

Combining all results, we get the formulation.

$$\lim_{m\to\infty}\frac{1}{m}\log E\{(\hat{p}_{ET})^2\} = -\frac{b^2}{\rho^2}\left(\frac{1+2\lambda^2}{1+\lambda^2}\right) \qquad \square$$

Note That Theorem 3.1 shows that the estimator is asymptotical optimal only in the case $\lambda \geq 0$. With $-\sqrt{1+\sqrt{2}} < \lambda < 0$, the second moment decreases faster than the first moment, but not twice as fast. For $\lambda \leq -\sqrt{1+\sqrt{2}}$, however, the second moment even decreases slower than the first moment. This result implies that sampling from the new measure is no more effective than general Monte Carlo.

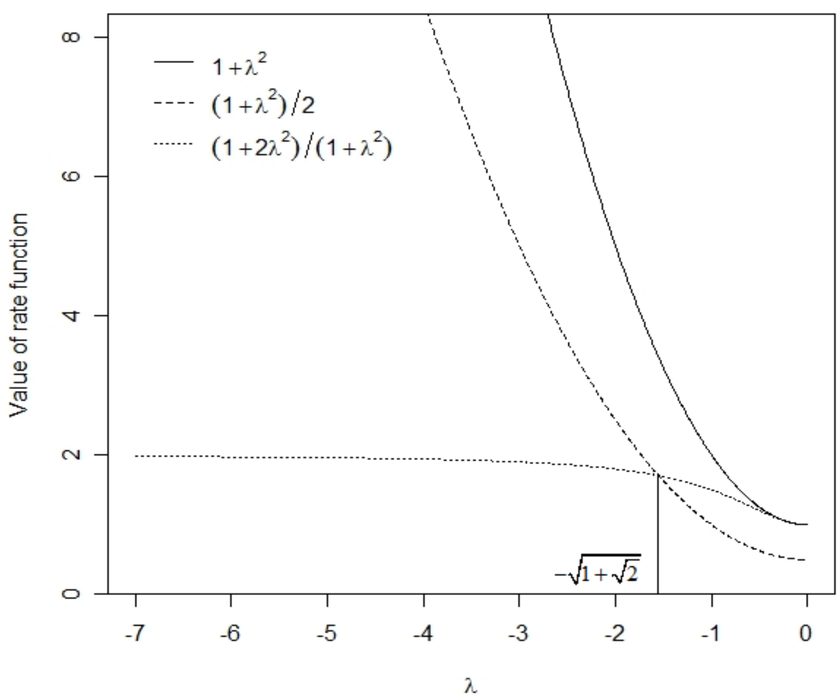

Fig. 3-4 *Comparison of rate function under* $\lambda < 0$ *for a single homogeneous model.*

In the case $\lambda < 0$, to eliminate the linear effect of $F_m(z)$, using exponential twist method with parameter $F_m^{'}(t_m)$ does not achieve the maximum utility. Review the properties of exponentially twisted procedure, we know that the maximal variance reduction occurs when parameter $F_m^{'}(t_m)$ makes the mean value of $f^*(z \mid \mu;\sigma;\lambda)$ locate at the point t_m. Consider $(\mu;\sigma;\lambda) = (0,1,\lambda)$ and $\lambda \geq 0$, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} z \frac{e^{t_m z}}{M_Z(t_m)} f(z \mid 0;1;\lambda)\, dz &= \frac{M_Z^{'}(t_m)}{M_Z(t_m)} \\
&= \frac{2t_m e^{\frac{t_m^2}{2}} \Phi(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m) + 2e^{\frac{t_m^2}{2}} \frac{\lambda}{\sqrt{1+\lambda^2}} \phi(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m)}{2e^{\frac{t_m^2}{2}} \Phi(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m)} \\
&= t_m + \frac{\lambda}{\sqrt{1+\lambda^2}} \frac{\phi(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m)}{\Phi(\frac{\lambda}{\sqrt{1+\lambda^2}} t_m)} \\
&\to t_m \quad \text{, if } \; m \to \infty
\end{aligned} \tag{3.6}$$

(3.6) show that the maximal utility will happen if we apply exponential twist method to factor. The phenomenon coincides with that GL suggest. However, if $\lambda < 0$, the mean value of $f^*(z \mid 0;1;\lambda)$ is

$$\int_{-\infty}^{\infty} z \frac{e^{t_m z}}{M_Z(t_m)} f(z \mid 0;1;\lambda)\, dz = \frac{M_Z^{'}(t_m)}{M_Z(t_m)}$$

$$= t_m + \frac{\lambda}{\sqrt{1+\lambda^2}} \frac{\phi(\frac{|\lambda|}{\sqrt{1+\lambda^2}} t_m)}{1-\Phi(\frac{|\lambda|}{\sqrt{1+\lambda^2}} t_m)}$$

$$\approx \frac{1}{1+\lambda^2} t_m \tag{3.7}$$

This equation holds because of $\phi(x)/1-\Phi(x) \sim x$. Observe (3.7), we know that the algorithm is less efficient if λ is getting smaller. This result in (3.7) also corresponds with Fig. 3-4. Namely, negative λ incurs a width which makes the variance increase. In the next Chapter, we will tailor the algorithm to eliminate the effect from shape parameter λ.

Chapter 4 The New Method for SN Factor

In Chapter 3, we know that asymptotical efficiency can not be achieved because the nonlinear behavior of $F_m(Z)$ is non-negligible. This suggests that to obtain further variance reduction we need to address the other component of $F_m(Z)$. Glasserman et al (1999) attempted to use stratification technique to decrease variability except for linear part. Here, we completely limit ourselves to IS methodology to build an efficient algorithm. Our approach emphasizes on choosing density "form" of Z rather than shifting, scaling or exponentially twisting. In the next section, we begin with a more general result in Chiang, Yueh, and Hsie (2007) (henceforth CYH) but for a different model.

4.1 Extension of CYH Importance Sampling Algorithm

The key idea in CYH is to find a simple alternative characterization of default event. To motivate the algorithm we take, observe the following proposition:

Proposition 4.1

Consider a single factor model where $c_i = 1$, $b_i = b$ and $\rho_i = \rho$; Random variables Z and ε_i follow $SN(0,1,\lambda)$ and $SN(0,1,0)$ respectively. Then the set $\{L_m > mq\}$ is equivalent to the event $\{Z > H_{[mq+1]}\}$ if $H_{[mq+1]}$ is denoted as $[mq+1]\,th$ order statistics of $\{H_i\}_{i=1}^m$, where

$$H_i = \frac{b\sqrt{m} - \sqrt{1-\rho^2}\,\varepsilon_i}{\rho}$$

Proof: Since

$$
\begin{aligned}
&I\{L_m > mq\} = 1 \\
&\Leftrightarrow \sum_i I\{X_i > b\sqrt{m}\} > mq \\
&\Leftrightarrow \sum_i I\{Z > \frac{b\sqrt{m} - \sqrt{1-\rho^2}\varepsilon_i}{\rho}\} > mq \\
&\Leftrightarrow I\{Z > H_{[mq+1]}\} = 1
\end{aligned}
$$

Hence, the event $\{L_m > mq\}$ is equivalent to the event $\{Z > H_{[mq+1]}\}$. □

Proposition 4.1 indicates a simple alternative characterization for the event $\{L_m > mq\}$. It provides a simpler way to ensure that for every replication where the set we interest always takes place. By Proposition 4.1, we create an estimator of single SN factor model as following

$$
I\{L_m > mq\}L_r \tag{4.1}
$$

where $L_r = 1 - F_Z(H_{[mq+1]})$ denotes the likelihood ratio and F_Z is the cumulative density function of Z. Clearly, (4.1) is not restricted to what the distribution of Z is. This means that the algorithm is allowed to general case. We will consider behavior of (4.1) in the following theorem. By analyzing the asymptotical performance, we can find a useful guideline for choosing appropriate IS density of Z.

Theorem 4.1

Consider a single factor model where $c_i = 1$, $b_i = b$, $\rho_i = \rho$ and (Z, ε_i) follow the same distribution assumption in Proposition 1, then (4.1) has bounded relative error.

Proof: First, we exploit the result shown in Lucas et al (2003). Assume that S_j is a latent variable which obeys the general factor model

$$S_j = g(f, \varepsilon_j)$$

where f is common factor, ε_j is specific risk factor, and $g(\cdot,\cdot)$ defines the functional form of the factor model. Lucas et al (2003) used the Theorem 12.13 of Williams (1991) and indicate that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} I\{S_j < s^*\} \xrightarrow{a.s} P(S_j < s^* \mid f)$$

By the same argument as Lucas et al (2003), we then have

$$\lim_{m\to\infty} \frac{L_m}{m} \xrightarrow{a.s} E[I\{X_i > b\sqrt{m}\} \mid Z]$$

So, we have

$$E[I\{L_m > mq\}] \to P(E[I\{X_i > b\sqrt{m}\} \mid Z] > q)$$

$$= P(Z > \frac{b\sqrt{m} - \Phi^{-1}(1-q)\sqrt{1-\rho^2}}{\rho})$$

$$= 1 - F_Z(\frac{b\sqrt{m} - \Phi^{-1}(1-q)\sqrt{1-\rho^2}}{\rho})$$

The second moment of (4.1) is written as (by Theorem 1.10 of Shao (1998) and Theorem 4.3.1 of Sen and Singer (1993))

$$E[I\{L_m > mq\}L_r^2]$$

$$= E[I\{L_m > mq\}(1 - F_Z((H_i)_{[mq+1]}))^2]$$

$$= E\left[I\{L_m > mq\}(1 - F_Z(\frac{b\sqrt{m} - \sqrt{1-\rho^2}(\varepsilon_i)_{m-[mq]}}{\rho})^2 \right]$$

$$\le (1 - F_Z(\frac{b\sqrt{m} - \Phi^{-1}(1-q)\sqrt{1-\rho^2} + o(1)}{\rho}))^2$$

Therefore, we have

$$\limsup_{m\to\infty} \frac{\sqrt{Var(I\{L_m > mq\}L_r)}}{E[I\{L_m > mq\}]}$$

$$= \limsup_{m\to\infty} \frac{\sqrt{E[I\{L_m > mq\}L_r^2] - E^2[I\{L_m > mq\}L_r]}}{E[I\{L_m > mq\}]}$$

$$< \infty$$ □

The method works well for portfolio whose tail behavior is dominated by a “key” random variable. To show that the results have content, we give two specific examples. For the first example, consider the model

$$X_i = \sqrt{\frac{\nu}{S}}(\rho_i Z + \sqrt{1-\rho_i^2}\,\varepsilon_i) \qquad , \quad i = 1, \ldots, m$$

Here Z is a standard normal random variable, ε_i are i.i.d standard normal random variables, and S is chi-square distribution with r degrees of freedom. Applying the key idea behind Proposition 1, we can find the set $\{S < (H_i)_{[mq+1]}\}$ where $H_i = (\rho_i Z + \sqrt{1-\rho_i^2}\,\varepsilon_i)/b\sqrt{m}$. It is the simpler expression of $\{L_m > mq\}$. Note that the sample of Z is generated from the original distribution. By a analogous algorithm in section 3.3, we get a efficient estimation of $\{L_m > mq\}$. In this model, the random variable S plays a key role to vanish variability. Changing the measure of S is sufficient to achieve substantial variance reduction.

Table 4-1：Comparison of different methods for $m = 250$; $\nu = 12$; $q = 0.25$.

$P(L_m > mq)$					
Method	**Algorithm 1 (Runs: 5×10^4)**		**CYH Method (Runs: 5×10^4)**		$V.R$
ρ	Prob. est	$S.E$	Prob. est	$S.E$	
0.1	8.58×10^{-6}	1.63×10^{-7}	8.53×10^{-6}	1.36×10^{-7}	1.43
0.2	9.74×10^{-6}	1.85×10^{-7}	9.75×10^{-6}	2.04×10^{-7}	0.82
0.3	1.18×10^{-5}	4.13×10^{-7}	1.18×10^{-5}	3.18×10^{-7}	1.68
0.4	1.39×10^{-5}	8.61×10^{-7}	1.42×10^{-5}	4.93×10^{-7}	3.05

Table 4-1 shows the performance of two estimators. Algorithm 1 is the suggestion of Bassamboo et al (2008) and we know that it has the bounded relative error. In the last column, we list the sample variance ratio $V.R$

$$V.R = \frac{[S.E(\widehat{p}_{A1})]^2}{[S.E(\widehat{p}_{CYH})]^2} \times \frac{5\times 10^4}{5\times 10^4}$$

, where $\widehat{p}_{A1}$ refers to the estimator of Algorithm 1 and $\widehat{p}_{CYH}$ refers to the estimator of CYH method. We find the fact that $\widehat{p}_{A1}$ and $\widehat{p}_{CYH}$ have analogous performance of simulation in Table 4-1. But, note that the implementation of the new method is more easily.

The next example illustrates the normal case discussed in Glasserman (2004). All obligors are divided into two blocks. The first block consists of m_1 obligors whose marginal default probability is p_1. This block is dominated by the factor Z_1 and has a common loading a_1. The second block comprises the last $m - m_1$ obligors. All obligors in the second block have marginal default probability p_2 and affected only by factor Z_2 with a common loading a_2. This model is

$$X_i = a_1 Z_1 + \sqrt{1 - a_1^2}\,\varepsilon_i, \quad i = 1, \ldots, m_1$$

$$X_j = a_2 Z_2 + \sqrt{1 - a_2^2}\,\varepsilon_j, \quad j = m_1 + 1, \ldots, m$$

Set $N_1 = \sum_{i=1}^{m_1} I\{X_i > \Phi^{-1}(1-p_1)\}$ and $N_2 = \sum_{j=m_1+1}^{m} I\{X_j > \Phi^{-1}(1-p_2)\}$, the equivalent set then is written as $\bigcup_i E_{i,\ell}$ where $E_{i,\ell} = \{(z_1, z_2) \mid N_1 + N_2 \ge \ell, N_1 = i\}$.

Table 4-2 shows the performance of general Monte Carlo and the CYH method. Note that variance reduction is measured relative to general Monte Carlo simulation. The CYH method provides an excellent performance than general Monte Carlo. The behavior of loss distribution seems to be successfully captured by the CYH method.

Table 4-2：Comparison for $(m, m_1; a_1; a_2; p_1; p_2) = (1{,}000; 150; 0.8; 0.7; 0.05; 0.001)$

$P(L>\ell)$					
ℓ	**Monte Carlo (Run 10^5)**		**CYH Method (Run 10^3)**		$V.R$
	Estimation	S.E	Estimation	S.E	
90	1.36×10^{-2}	3.66×10^{-4}	1.41×10^{-2}	8.26×10^{-5}	1965
110	6.71×10^{-3}	2.57×10^{-4}	6.99×10^{-3}	4.65×10^{-5}	3073
130	2.63×10^{-3}	1.61×10^{-4}	2.69×10^{-3}	2.11×10^{-5}	5947
150	3.80×10^{-4}	6.16×10^{-5}	4.46×10^{-4}	2.92×10^{-6}	44261

Although using order statistic increases simplicity, the flexibility is restricted simultaneously. For instance, if all the exposures c_i are different from each other, then the sorting and partitioning procedures make the method time consuming. Obviously, the original problem in estimating rare event is transferred into another one. For more discussions about determination of key random variable is referred in Lucas et al (2003).

4.2 The Proposed algorithm for Skew Factor Model

Note the conclusion in Theorem 4.1, we know the likelihood L_r has an excellent utility in variance reduction. Although the new method is inflexible to tackle inhomogeneous portfolio, it provides a way to build IS density for Z. In the following, we will introduce the strategy to search an effective IS algorithm.

Consider the result described in section 3.3, we know that vanishing linear variability of $F_m(z)$ can increase the efficiency of simulation except for $\lambda < 0$. Therefore, the procedure of eliminating the linear part of $F_m(z)$ is essential. This means the new likelihood ratio $L_r^{New}(z) = f(z \mid 0;1;\lambda) / f_{New}(z)$ must contain the function $\exp(-t_m z)$, namely

$$L_r^{New}(z) \propto \exp(-t_m z) \tag{4.2}$$

Furthermore, in the second part of Theorem 3.1, we find that the nonlinear behavior of $F_m(z)$ seriously effect the efficiency of variance reduction. To eliminate this effect from the nonlinear part, we consider the limit regime rather than integral itself. We focus on modifying the other part of density of Z but for vanishing nonlinear part of $F_m(z)$ directly. With the definition of asymptotically optimal, we need to find a likelihood ratio which decrease as fast as possible if we apply IS to Z.

In Theorem 4.1, L_r is of the bounded relative property. It is reasonable to utilize asymptotical decay rate of L_r to create appropriate IS density. In multifactor and inhomogeneity case, however, getting a likelihood ratio like L_r is difficult. So we turn attention to setting where expectation of likelihood ratio decays in the same rate of L_r. In other words, we expect the following equation holds

$$\lim_{m\to\infty} \frac{\log L_r}{\log E[L_r^{New}(z)]} = 1 \tag{4.3}$$

Once we find a new $L_r^{New}(z)$ satisfying (4.2) and (4.3), the corresponding IS density is then determined. Note that the combinative way leads to not only vanishing the linear effect but considering the nonlinear part of $F_m(z)$ simultaneously.

To represent our procedure precisely, we consider the setting where $Z \sim SN(0,1,\lambda)$ and $\lambda < 0$. Clearly, it is difficult to directly calculate

$$L_r = \int_{\left(\frac{b\sqrt{m}-\sqrt{1-\rho^2}\varepsilon_i}{\rho}\right)_{[mq+1]}}^{\infty} 2\phi(t)\Phi(\lambda t)dt$$

If m is large sufficiently, an approximation for the integral (e.g, Shao 1998, Chap. 1 and Sen and Singer 1993, Chap. 4) suggests that

$$L_r \sim \int_{\frac{b\sqrt{m}+\Phi^{-1}(q)\sqrt{1-\rho^2}}{\rho}}^{\infty} 2\phi(t)\Phi(\lambda t)dt$$

Therefore, for any small value δ, L_r is simplified into

$$L_r = 2\delta\phi(t_m + o(\sqrt{m}))\Phi(\lambda t_m + o(\sqrt{m}))$$

The last equation holds because of $t_m = O(\sqrt{m})$. This discussion of t_m is shown in GL. Then, the associated new IS density f_{New}^{N} is written as

$$f_{New}^{N}(z) = \phi(z - t_m) \tag{4.4}$$

Note that the choice of IS density makes $E[L_r^{New}(z)]$ satisfy (4.3), that is

$$\begin{aligned} E[L_r^{New}(z)] &= 2\int_{-\infty}^{\infty} \frac{\phi(z)\Phi(\lambda z)}{\phi(z - t_m)}\phi(z - t_m)dz \\ &= 2\int_{-\infty}^{\infty} \exp(-t_m z + \frac{t_m^2}{2})\Phi(\lambda z)\phi(z - t_m)dz \\ &= 2\exp(-t_m(\xi + t_m) + \frac{t_m^2}{2})\ \Phi(\lambda(\xi + t_m)) \\ &\sim L_r \end{aligned}$$

where ξ denotes a constant. The last equation holds because of the second mean value theorem for integral. Note that the choice of an appropriate IS density in such procedure is not unique. For instance, the following density

$$f_{New}^{N}(z) = 2\phi(z - t_m)\Phi(|\lambda|(z + t_m))$$

is another feasible one. By the similar argument, a appropriate IS density for $\lambda \geq 0$ is

$$f_{New}^{P}(z) = \frac{2}{M(t \mid 0;1;\lambda)} e^{t_m z} \phi(z) \Phi(\lambda z) \tag{4.5}$$

Especially, when $\lambda = 0$, f_{New}^{P} becomes the IS density GL suggest. For notational simplicity, considering $(\mu, \sigma) = (0,1)$, we build our algorithm as following:

1. Calculate $t_m = \arg\max_t \{F_m(t) - t^2 / 2\}$.
2. Check value λ of Z, choose (4.5) as IS density of Z if $\lambda \geq 0$ and (4.4) otherwise.
3. Set t_m to the IS density.
4. Sampling Z and calculate the product $L_r^{New}(z)$ of each likelihood ratio.
5. Compute $\theta_m(z;q)$.
6. Return the estimate $I\{L_m > mq\} \tilde{L}_r$

where $\tilde{L}_r = e^{-\theta_m(Z;q)L_m + \psi_m(\theta_m(Z;q),Z)} L_r^{New}$ is the combined likelihood ratio. If we repeat step 1 to 6 .˙ times, an estimator $\hat{p}_{New}$ can be constructed by averaging the .˙ values of the estimates, we have a estimation for $P(L_m > mq)$ under skew normal copula model. Once we have selected a new parameter vector t_m which satisfies

$$t_m = \arg\max_t \{F_m(t) - \frac{1}{2} t^T t\},$$

choosing (4.4) or (4.5) and component of t_m for single IS density; we can easily extend the single factor to multiple factors.

4.3 Asymptotic Optimality

We now consider the performance of the estimator $\hat{p}_{New}$. The strength of our proposed lies in its variance reduction efficiency established by the following theorem:

Theorem 4.2

Consider the same assumption in Theorem 4.1, then

(a) For $\lambda \geq 0$

$$\lim_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] = -\frac{b^2}{2\rho^2}$$

$$\lim_{m\to\infty} \frac{1}{m} \log E[(\hat{p}_{New})^2] = -\frac{b^2}{\rho^2}$$

(b) For $\lambda < 0$

$$\lim_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] = -\frac{b^2(1+\lambda^2)}{2\rho^2}$$

$$\lim_{m\to\infty} \frac{1}{m} \log E[(\hat{p}_{New})^2] = -\frac{b^2(1+\lambda^2)}{\rho^2}$$

Proof: For $\lambda \geq 0$, the proof is the same as (a) in Theorem 3.1. We consider the part (b) directly.

(b): First we show that

$$\liminf_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] = -\frac{b^2(1+\lambda^2)}{2\rho^2}$$

By the similar argument in Theorem 3.1, we know for arbitrary $\tau > 0$

$$E[I\{L_m > mq\}] = P(L_m > mq \mid p(Z) = q + \upsilon) P(p(Z) > q + \upsilon)$$

$$\geq \frac{1}{2} P(Z > \frac{b\sqrt{m} + \Phi^{-1}(q+\upsilon)}{\rho})$$

$$\geq \phi(z_m + \tau)\Phi(\lambda(z_m + \tau))$$

where $z_m = \{b\sqrt{m} + \Phi^{-1}(q+\upsilon)\}/\rho$. We have

$$\liminf_{m\to\infty} \frac{1}{m} \log E[I\{L_m > mq\}] \geq \liminf_{m\to\infty} \frac{1}{m} \log \phi(z_m + \tau)$$

$$+ \liminf_{m\to\infty} \frac{1}{m} \log \Phi(\lambda(z_m + \tau))$$

Note that

$$\liminf_{m\to\infty} \frac{1}{m} \log \phi(z_m + \tau) = -\frac{b^2}{2\rho^2}$$

Applying the property $\phi(x)/\Phi(-x) \sim x$ as $x \to \infty$, we get

$$\liminf_{m\to\infty} \frac{1}{m} \log \Phi(\lambda(z_m + \tau)) = \liminf_{m\to\infty} \frac{-b|\lambda|\phi(\lambda(z_m+\tau))}{2\rho\Phi(\lambda(z_m+\tau))}$$

$$= \liminf_{m\to\infty} \frac{-b|\lambda|\phi(|\lambda|(z_m+\tau))}{2\rho\sqrt{m}\Phi(-|\lambda|(z_m+\tau))}$$

$$= \liminf_{m\to\infty} \frac{-b|\lambda|}{2\rho\sqrt{m}} \{|\lambda|(z_m+\tau) + o(\sqrt{m})\}$$

$$= -\frac{b\lambda^2}{2\rho^2}$$

Next, we show

$$\limsup_{m\to\infty} \frac{1}{m} \log E[(\hat{p}_{New})^2] \le -\frac{b^2(1+\lambda^2)}{\rho^2}$$

By the similar discussion in Theorem 3.1, we have

$$\begin{aligned} E[(\hat{p}_{New})^2] &= E[I\{L_m > mq\}\tilde{L}_r^2] \\ &\le 4E[\exp\{2F_m(Z) - t_m Z + t_m^2\}\Phi^2(\lambda Z)] \\ &\le 4\exp\{2F_m(t_m) - t_m^2\}E[\Phi^2(\lambda Z)] \\ &\le 4\exp\{2F_m(t_m) - t_m^2\}E[\Phi^2(\lambda Z + \lambda t_m)] \end{aligned}$$

Using the second mean value theorem for integral, for a small value ζ, we get

$$\begin{aligned} \limsup_{m\to\infty} \frac{1}{m} \log E[(\hat{p}_{New})^2] &\le \limsup_{m\to\infty} \frac{1}{m}\{-z_m^2(1-\zeta)\} \\ &\quad + \limsup_{m\to\infty} \frac{1}{m} \log E[\Phi^2(\lambda Z + \lambda t_m)] \\ &= -\frac{b^2}{\rho^2} + \limsup_{m\to\infty} \frac{1}{m} \log \Phi^2(\lambda t_m + o(\sqrt{m})) \end{aligned}$$

Observe that

$$\limsup_{m\to\infty}\frac{1}{m}\log\Phi^2(\lambda t_m+o(\sqrt{m}))=\limsup_{m\to\infty}\frac{2}{m}\log\Phi(\lambda z_m+o(\sqrt{m}))$$

$$=\limsup_{m\to\infty}\frac{-b|\lambda|}{\rho\sqrt{m}}\{|\lambda|z_m+o(\sqrt{m})\}$$

$$=-\frac{b^2\lambda^2}{\rho^2}$$

Combining all the result and applying Jensen's inequality we complete the proof. □

This result indicates that our proposed IS algorithm should be effective in estimating loss distribution. Even though the assumption in Theorem 4.2 is for homogeneous single factor model, the proposed algorithm is practicably applied to multifactor and inhomogeneity cases. Note that our proposed algorithm does not require what density the factor Z should follow. When the specific factors are of arbitrary distribution, we need only to modify the associated $f_{New}(z)$ to satisfy equation (4.3). In next chapter, our numerical results for skew normal factor model also confirm the expectation.

Chapter 5 Implementation Issues

In this chapter we compare performance of the new estimator $\hat{p}_{New}$ with general Monte Carlo simulation. We investigate sensitivity to λ, b, ρ and q. The broad conclusions are that the new algorithm provides significant improvement over the performance of general Monte Carlo simulation. This improvement increase as the event becomes rare. This result supports our theoretical conclusions that the sample variance ratio, as measured by the ratio of the standard deviation of general Monte Carlo simulation to the standard deviation of $\hat{p}_{New}$, remains well behaved as the probability of large losses becomes increasingly rare.

For implementation of new algorithm, (4.4) is easily generated using the inverse transform method. However, the cumulative distribution associated with (4.5) does not have a closed form. It is not straightforward to use the inverse transform methods to generate samples from this distribution. Instead, we use a root-finding method of numerical integration to generate samples we need.

Our first example is a single factor portfolio of $m = 1{,}000$ and $Z \sim SN(0,1,\lambda)$. The model parameters are chosen to be $q = 0.4$, $b = 0.0345$, $\rho = 0.3$ and exposure $c_i = 1$. We generate 5,000 samples for proposed algorithm and 100,000 samples for general Monte Carlo simulation. Table 5-1 reports samples variance ratio for several values of λ in estimating $P(L_m > mq)$.

Table 5-1: Variance Reduction for decreasing λ.

$P(L_m > mq)$					
Method	**General Monte Carlo (Runs:** 1×10^5 **)**		$\hat{p}_{New}$ **(Runs:** 5×10^3 **)**		$V.R$
λ	Prob. est	$S.E$	Prob. est	$S.E$	
1.0	4.71×10^{-3}	2.17×10^{-4}	4.82×10^{-3}	1.15×10^{-4}	65
0.5	4.46×10^{-3}	2.10×10^{-4}	4.63×10^{-3}	1.03×10^{-4}	82
-0.5	2.70×10^{-4}	5.19×10^{-5}	3.04×10^{-4}	8.49×10^{-6}	748
-1.0	1.00×10^{-5}	9.98×10^{-6}	9.42×10^{-6}	3.49×10^{-7}	16281

At small value of λ, the variance ratio becomes very large. The performance of $\hat{p}_{New}$ is significantly better than general Monte Carlo simulation. The improvement is substantial especially for negative value of λ. Note that the variance ratio rapidly changes when negative value λ varies slowly.

Table 5-2: Variance Reduction for increasing b.

$P(L_m > mq)$					
Method	**General Monte Carlo (Runs:** 1×10^5 **)**		$\hat{p}_{New}$ **(Runs:** 5×10^3 **)**		$V.R$
b	Prob. est	$S.E$	Prob. est	$S.E$	
0.0345	4.71×10^{-3}	2.17×10^{-4}	4.82×10^{-3}	1.15×10^{-4}	65
0.0375	1.62×10^{-3}	1.27×10^{-4}	1.76×10^{-3}	4.53×10^{-5}	153
0.0405	5.61×10^{-4}	7.48×10^{-5}	5.65×10^{-4}	1.54×10^{-5}	469
0.0435	1.50×10^{-4}	3.87×10^{-5}	1.62×10^{-4}	4.62×10^{-6}	1399

Table 5-2 shows the performance of the proposed algorithm as b changes. Again we set $m = 1{,}000$, $q = 0.4$, $\lambda = 1$. The factor loading ρ is kept fixed at 0.3, each $c_i = 1$. We generate 5,000 samples for proposed algorithm and 100,000 samples for general Monte Carlo simulation. In last column, we observe that all performances are significantly better than general Monte Carlo simulation. The variance ratio improves as b increases.

Table 5-3 shows performance of the proposed algorithm as factor loading ρ changes. In this case, the parameters of model are $m = 1{,}000$, $q = 0.4$, $\lambda = -1$, $b = 0.0345$ and $c_i = 1$. We generate 5,000 samples for proposed algorithm and 100,000 samples for general Monte Carlo simulation. All results perform significantly better than general Monte Carlo simulation, especially when ρ decrease.

Table 5-3: Variance Reduction for increasing ρ.

$P(L_m > mq)$					
Method	**General Monte Carlo (Runs:** 1×10^5**)**		$\hat{p}_{New}$ **(Runs:** 5×10^3**)**		$V.R$
ρ	Prob. est	$S.E$	Prob. est	$S.E$	
0.3	1.00×10^{-5}	9.98×10^{-6}	9.42×10^{-6}	3.49×10^{-7}	16281
0.35	1.01×10^{-4}	3.16×10^{-5}	6.16×10^{-5}	2.05×10^{-6}	4756
0.4	3.10×10^{-4}	5.56×10^{-5}	2.73×10^{-4}	8.58×10^{-6}	840
0.45	7.01×10^{-4}	8.36×10^{-5}	7.76×10^{-4}	2.41×10^{-5}	241

Our next example is a multifactor portfolio of $m = 1{,}000$ and $Z_j \sim SN(0,1,\lambda_j)$, $j = 1,\cdots,5$. Each factor Z_j has shape parameter λ_j which is generated uniformly from the interval (-1,0) The exposures c_i is kept fixed at 1; b_i and a_{ij} are distributed uniformly from (0.02, 0.07) and $(0, 1/\sqrt{5})$ respectively. Table 5-4 compares the performance of the proposed algorithm with general Monte Carlo simulation as q change. The general Monte Carlo simulation results are based on 50,000 replications whereas the number of IS replications is 1,000. When the loss level is small, the proposed algorithm is a bit better than general Monte Carlo simulation. At large values of q, the $\{L_m > mq\}$ becomes rare and then the variance ratio becomes large. The improvement is obvious for q in the range of 0.45 to 0.5.

Table 5-4: Variance Reduction for increasing q.

$P(L_m > mq)$					
Method	**General Monte Carlo (Runs:** 5×10^5 **)**		$\hat{p}_{New}$ **(Runs:** 1×10^3 **)**		$V.R$
q	Prob. est	$S.E$	Prob. est	$S.E$	
0.35	1.90×10^{-3}	1.94×10^{-4}	1.91×10^{-3}	1.79×10^{-4}	59
0.40	8.01×10^{-4}	1.26×10^{-4}	7.55×10^{-4}	7.22×10^{-5}	153
0.45	2.20×10^{-4}	6.63×10^{-5}	2.72×10^{-4}	2.59×10^{-5}	325
0.50	1.00×10^{-4}	4.47×10^{-5}	1.00×10^{-4}	1.09×10^{-5}	827

Chapter 6 Concluding Remarks

In this thesis, we have proposed a new algorithm for estimation of tail probability in skew normal copula model. We started with the case of applying exponential twist technique to the default random variables conditional on common factors. However, GL show that the conditional IS estimator does not achieve asymptotical optimal unless the correlation between obligors is very weak. Therefore, GL further suggest shift the mean of underlying factor to eliminate the residual variability. This procedure makes the algorithm asymptotical optimal.

Different from the normal copula model, however, the leptokurtic and asymmetric characters of skew normal result in the situation where second moment of IS estimator converge in unintelligible decreasing rate. So, to choose IS density of underlying factors becomes intricate. To improve the efficiency of simulation, we intuitively consider the usual exponential twist to eliminate the linear part of $F_m(z)$. Surprisingly, using exponential twist does not guarantee variance reduction. A way to speed up the decreasing rate of likelihood ratio is necessary.

Further analyze the failure of case (3.7), we know that the achievement of optimality depends on the location of mean value. Once the mean value locate on a specific point t_m, we will obtain the maximal utility of simulation. We had considered a new exponential twist density $e^{H(z)} f(z|0;1;\lambda)/M_Z(H(z))$ where $H(t_m) = F_m^{'}(t_m)$ and $H(t_m)$ simultaneously satisfy the following equation

$$\int_{-\infty}^{\infty} z \frac{e^{H(t_m)z}}{M_Z(H(t_m))} f(z \mid 0;1;\lambda)\, dz \to t_m$$

However, searching the function $H(\cdot)$ is not easy in practice and that is to be the one direction of future work. We next extend the CYH method to solve utility problem. By finding an asymptotical behavior, we can decide the IS density of factors.

Note that our proposed algorithm is also applied to other factor assumption. Because our consideration of building IS density put emphasis on adjusting the width of a distribution to mimic the form of optimal density but for the determination of the optimal shifting. We have successfully extended single factor assumption in CYH to multifactor cases and illustrated its effectiveness in more complex cases through numerical results. The other direction of future work is to extend the approach to different factor assumption.

REFERENCE

[1] Anderson, L, and J. Sidenius., " Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings." *Journal of Credit Risk*, 1, No.1, pp. 29-70, 2005.

[2] Arnold, B. C, and G. D. Lin, "Characterizations of the Skew-Normal and Generalized Chi Distribution." *Sankhyā : The Indian Journal of Statistics*, 66, No.4, pp. 593-606, 2004.

[3] Azzalini, A, "A Class of Distributions Which Include the Normal Ones." *Scandinavian Journal of Statistics*, 12, pp. 171-178, 1985.

[4] Bassamboo, A, S. Juneja, and A. Zeevi, "Portfolio Credit Risk with Extremal Dependence: Asymptotic Analysis and Efficient Simulation." *Operations Reseach,* 56, pp. 593-606, 2008.

[5] Bluhm, C, L. Overbeck, and C. Wagner, "An Introduction to Credit Risk Modeling." *CRC Press/Chapman & Hall*, 2002.

[6] Bucklew, J. A, *Large Deviation Techniques in Decision, Simulation, and Estimation*, John Wiley, New York, 1990.

[7] Capriotti, L, "Least-Square Importance Sampling for Monte Carlo Security Pricing." *Quantitative Finance*, 8, No.5, pp 485-497, 2008.

[8] Chiang, M. H, M. L. Yueh, and M.H. Hsieh, "An Efficient Algorithm for Basket Default Swap Valuation." *Journal of Derivatives,* pp. 8-19, 2007.

[9] Crouhy, M, D. Galai, and R. Mark, "A Comparative Analysis of Current Credit Risk Models." *Journal of Banking and Finance*, 24, pp. 59-117, 2000.

[10] Glasserman, P, "Tail Approximations for Portfolio Credit Risk." *Journal of Derivatives*, 12, No. 2, pp. 24-42, 2004.

[11] Glasserman, P, and J. Li, “Importance Sampling for Portfolio Credit Risk.” *Management Science*, 55, No.11, pp. 1643-1656, 2005.

[12] Glasserman, P, P. Heidelberger, and P. Shahabuddin, "Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options." *Mathematical Finance*. 9, No. 2, pp. 117-152, 1999.

[13] Glasserman, P, W. Kang, and P. Shahabuddin, "Fast Simulation of Multifactor Portfolio Credit Risk." *Operations Research.* 56, No.5, pp. 1200-1217, 2008.

[14] Graciela González-Farías., D. M. Armando, A. K. Gupta, “Additive Properties of Skew Normal Random Vectors.” *Journal of Statistical Planning and Inference*. 126, pp. 521-534, 2004.

[15] Gupta, A. K, T. T. Nguyen, and J. A. Sanqui, “Characterization of the Skew-Normal Distribution.” *Annals of the Institute of Statistical Mathematics*, 56, pp. 351-360, 2004.

[16] Gupta, G, C. Finger, M. Bhatia. *CreditMetrics Technical Document*. J.P.Morgan ＆ CO, New York, 1997.

[17] Hull, J, and A. White, “Valuation of CDO and nth to Default CDS without Monte Carlo Simulation.” *Journal of Derivarives*, 12, No. 2, pp. 8-23, 2004.

[18] Johnson, N. L, S. Kotz, A. W. Kemp, *Univariate Discrete Distributions*, 2^{nd} ed. John Wiley and Sons, New York, 1992.

[19] Kostadinov, K, “Tail Approximation for Credit Risk Portfolios with Heavy-Tailed Risk Factors.” *Technique Report*, The Munich University of Technology, 2005.

[20] Li, D,”On Default Correlation: A Copula Function Approach.” *Journal of Fixed Income*, 9, pp. 43-54, 2000.

[21] Lucas, A, P. Klaassen, P. Spreij, and S. Straetmans, “Tail Behavior of Credit Loss Distributions for General Latent Factor Models.“ *Applied Mathematical Finance.*

10, No. 4, pp. 337-357, 2003.

[22] McNeil, A.J, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concept, Techniques, and Tools*. Princeton University Press, 2005.

[23] Ney, P, "Dominating Points and the Asymptotics of the Large Deviations for Random Walk on R^d." *Annals of Probability*. 11, No. 1, pp. 158-167, 1983.

[24] O'Hagan and T. Leonard, "Bayes Estimation Subject to Uncertainty about Parameter Constraints." *Biometrika*, 63, pp. 201-202, 1976.

[25] Sadowsky, J. S., J. A. Bucklew, "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation." *IEEE Transactions on Information Theory*, 36, No.3, May, pp. 579-588, 1990.

[26] Schloegl, L, and D. O'Kane, "A Note on the Large Homogenous Portfolio Approximation with the Student-t Copula." *Finance and Stochastics*, 9, pp. 577-584, 2005.

[27] Sen, P. K, J. M. Singer, *Large Sample Methods in Statistics*. Chapman & Hall, London, 1993.

[28] Shao, J, *Mathematical Statistics*. Spring-Verlag, New York, 1998.

[29] Williams,D, *Probability with Martingales*. Cambridge: Cambridge University Press, 1991.